AD-766 870

THE SENSITIVITY AND COMPARISON OF GENERAL ATMOSPHERIC CIRCULATION MODELS

Michael Warshaw

RAND Corporation
Santa Monica, California

November 1972

THE SENSITIVITY AND COMPARISON OF GENERAL ATMOSPHERIC CIRCULATION MODELS

Michael Warshaw

November 1972

P-4933

15

/

November 1972

THE SENSITIVITY AND COMPARISON OF GENERAL ATMOSPHERIC CIRCULATION MODELS

Michael Warshaw

The Rand Corporation, Santa Monica, California

INTRODUCTION

The use of a computer simulation of the global atmospheric circulation is the most promising way of answering the climatological questions posed by the Climatic Impact Assessment Program. Among the many problems inherent in the use of such models, we have chosen to investigate the following three:

- How sensitive is the model to changes in boundary conditions? That is, can an observable change in a predicted atmosphere be causally related to the boundary condition change or is it just an acceptable random state resulting from a slight perturbation or uncertainty in initial conditions?
- Models claiming to predict the same physical phenomena should agree if they begin with identical boundary and initial conditions. What is meant by "agreement" in the presence of small differences in initial conditions and do, in fact, existing models agree?
- A model, ideally, should predict the real atmosphere. Does this happen? Or alternatively, what actual comparisons are possible between a model and a very large file of real climatological data?

The eventual purpose of this study is to formulate statistical models, testing procedures and data retrieval programs which will help

answer these questions. This paper will describe some of our past and present approaches and experience with the first of the three tasks, model sensitivity, then briefly describe some future plans for tasks two and three. It is, in regards to present work, too soon for conclusive results. For this reason, we choose to maintain an informal, discursive tone in this paper, using a minimum of formalism, saving that for the time when definitive results are available.

THE ORIGINAL EXPERIMENT

In all the discussion to follow the word "model" refers to the Mintz-Arakawa 2-level general circulation model (Gates, et al., 1971; Arakawa, 1972). For some time, the almost canonical question (or scapegoat problem) was "what would be the effect on the climate if all the Arctic sea ice were replaced by water at -1°C (water temperatures in the M-A model are not permitted to vary). We entered the arena early in 1972 (Warshaw and Rapp, 1972) with the following experiment:

The initial conditions were chosen to simulate the state of the atmosphere for December 31 and the model run to simulate the next 60 days. This run is called the "control" run. Next, the initial global temperature field at both the σ = .25 and σ = .75 levels were perturbed with an additive random temperature. These random "noise" temperatures were drawn from a normal distribution having zero mean and a 1°C standard deviation. The model was re-run for the same 60 days as the control run. Following this, the original temperature field was again perturbed by *new* random additive noise drawn from the same distribution and the model run again for the same 60 days. We now have 3 runs, identical in all respects except for very small changes in the initial temperature field.

The experiment is completed by re-running each of the above 60-day simulations *except the Arctic sea ice is removed and replaced with sea water*. The initial temperature fields are, however, identical to runs 1-3.

Using these 6 runs, we formulate an analysis of variance model and proceed to test the hypothesis that removing the Arctic ice had no effect on the climatological variables of interest. The 2 × 3 layout is illustrated in Table 1, where Y_{ij} for the original experiment was the zonal average of the variable, averaged also over the last 30 days of the run. The procedure has been extended to a multivariant test, so that the Y_{ij} may now be vector valued, but Y_{ij} was only univariate in the reported work.

Table 1

ARRAY OF SCALAR VARIABLES

	Unperturbed	Perturbation No. 1	Perturbation No. 2	
Ice In	Y_{11}	Y_{12}	Y_{13}	$Y_{1.}$
Ice Out	Y_{21}	Y_{22}	Y_{23}	$Y_{2.}$
	$Y_{.1}$	$Y_{.2}$	$Y_{.3}$	

We may interpret differences in variables in the same *row* as "noise" in the experiment and differences in variables in the same column as useful "signal." An analysis of variance procedure provides a sharp procedure for testing whether the row means, $Y_{i.}$, are significantly different by eliminating the effects of differences in the column means, $Y_{.j}$. Table 2 reproduces the results given in Warshaw and Rapp (1972), and we see that many changes in temperature, geopotential height, zonal wind and heat transport may be statistically attributed to the removal of the Arctic sea ice.

Some important points remain to be made relative to this original experiment. First, it's readily apparent that the economics of this

Table 2

DATA AND STATISTICS FOR TEMPERATURE, GEOPOTENTIAL HEIGHT, ZONAL WIND, POLAR HEAT TRANSPORT, AND POLAR MOMENTUM TRANSPORT

Variable	Region	Altitude (mb)	Ice-In			Ice-Out			Mean Ice-In/Out Difference	F-Statistic (1,2) Deg Freedom	Significance
			Control	Pert. 1	Pert. 2	Control	Pert. 1	Pert. 2			
Temperature, T, °C	70-90°N	1000	-21.681	-23.188	-19.987	-6.044	-5.382	-6.145	-15.765	189	Reject @ .005
		800	-26.432	-26.624	-24.63	-17.284	-16.155	-17.342	-8.970	95	Reject @ .01
		400	-48.884	-50.279	-47.277	-53.084	-49.974	-52.166	2.929	3.23	Accept
	54-70°N	1000	-11.823	-11.121	-11.034	-6.561	-6.021	-6.086	-5.104	3169	Reject @ << .005
		800	-19.101	-18.916	-18.214	-14.218	-14.034	-14.008	-4.658	432	Reject @ << .005
		400	-45.866	-47.449	-45.835	-46.185	-46.142	-45.809	-0.339	4.67	Accept
	30-54°N	1000	4.184	4.703	4.072	4.525	4.606	4.84	-0.338	1.82	Accept
		800	-5.230	-4.745	-5.229	-1.703	-4.762	-4.472	-1.423	1.74	Accept
		400	-32.261	-32.172	-32.583	-33.353	-32.746	-32.785	0.622	5.82	Accept
Temperature Gradient ΔT, °C	86°N	T(1000) - T(400)	25.949	25.893	27.448	52.368	48.892	52.621	-24.869	519	Reject @ << .005
Geopotential Height H, Hectometers	70-90°N	1000	4.705	4.667	4.731	4.404	4.473	4.322	0.302	23.5	Reject @ .05
		800	20.951	20.944	21.126	21.441	21.532	21.445	-0.466	35.3	Reject @ .025
		400	68.647	68.491	6[illegible].202	69.393	69.840	69.472	-0.788	6.38	Accept
	54-70°N	1000	6.304	6.243	6.296	6.220	6.213	6.221	0.06	14.5	Accept
		800	22.970	22.937	23.039	23.279	23.291	23.285	-0.303	110	Reject @ .01
		400	71.008	70.765	71.195	71.739	71.737	71.771	-0.760	43.5	Reject @ .025
Geopotential Height Gradient, ΔH	H(86) - H(66)	800	-3.756	-3.592	-3.58	-3.492	-3.470	-3.565	-0.134	3.43	Accept
		400	-3.370	-2.814	-2.775	-3.230	-2.900	-3.481	0.217	0.74	Accept
	H(66) - H(46)	800	-1.322	-1.456	-1.199	-0.948	-0.990	-0.981	-0.353	23.7	Reject @ .05
		400	-3.517	-4.047	-3.110	-2.404	-2.669	-2.545	-1.019	18.1	Reject @ .05
Zonal Wind Velocity V, m sec^{-1}	74-82°N	800	-0.171	-1.937	0.220	0.805	1.347	1.576	-1.872	6.86	Accept
		400	3.708	0.896	2.[illegible]73	5.580	4.697	6.794	-3.265	21.6	Reject @ .05
	46-62°N	800	7.789	9.790	7.434	5.911	7.278	6.816	1.670	9.0	Accept
		400	18.639	21.458	17.508	14.[illegible]	16.282	15.580	3.856	15.3	Reject @ .05
Polar Heat Transport F, cal sec^{-1}	70°N	800	2.308×10^{14}	2.668×10^{14}	3.164×10^{14}	1.306×10^{14}	1.324×10^{14}	1.055×10^{14}	1.485×10^{14}	20.6	Reject @ .05
		400	1.012×10^{14}	-1.346×10^{14}	-1.146×10^{14}	0.190×10^{14}	0.770×10^{14}	-5.484×10^{14}	1.015×10^{14}	0.53	Accept
Polar Momentum Transport, M g cm^2 sec^{-2}	70°N	Total	1.265×10^{26}	1.443×10^{26}	1.246×10^{26}	0.771×10^{26}	0.639×10^{26}	0.977×10^{26}	0.522×10^{26}	11.3	Accept

procedure are totally unacceptable. The cost of simulating a single day by use of the Mintz-Arakawa model is approximately \$230,[†] thus the cost of replicating each run at least three times in order to provide adequate sampling is hardly the way of the future. An intermediate approach might be to run $n(\geq 3)$ replications of the control case each with some initial condition perturbation, but make only one run of the experiment which embodies the change in boundary conditions. For scalar variates, this is the standard t-test giving less sharp results than the F-test while only reducing the computing by 30 percent. Ideally we'd like to make only one control run and one experimental run. Below is an example of how to get into trouble with this approach, followed by a discussion of methods under study to extract more needed information from the sample provided by only one control and one experimental run.

THE CASE OF THE ALEUTIAN LOW

Figure 1 shows the location (latitude and longitude) and pressure of a local minimum in the North Pacific called the Aleutian low. There are six data points, 3 for the ice-in simulations and 3 for ice-out. Originally it was claimed that the removal of the Arctic ice *caused* a significant lowering of the pressure and westward shift of the low pressure center, as inferred from the two nonperturbed data points (Fletcher, 1971). Upon running the simulation four more times with the already described perturbations, it became clear (and you don't need a statistician to see it) that the shift and lowering of the pressure was well within the random variation of the triple of numbers (latitude, longitude, pressure).

This is exactly the sort of problem which will have to be faced over and over again when searching simulation results for significant change in the presence of the very high cost of additional samples.

[†] Using the IBM 360/91 computer at UCLA.

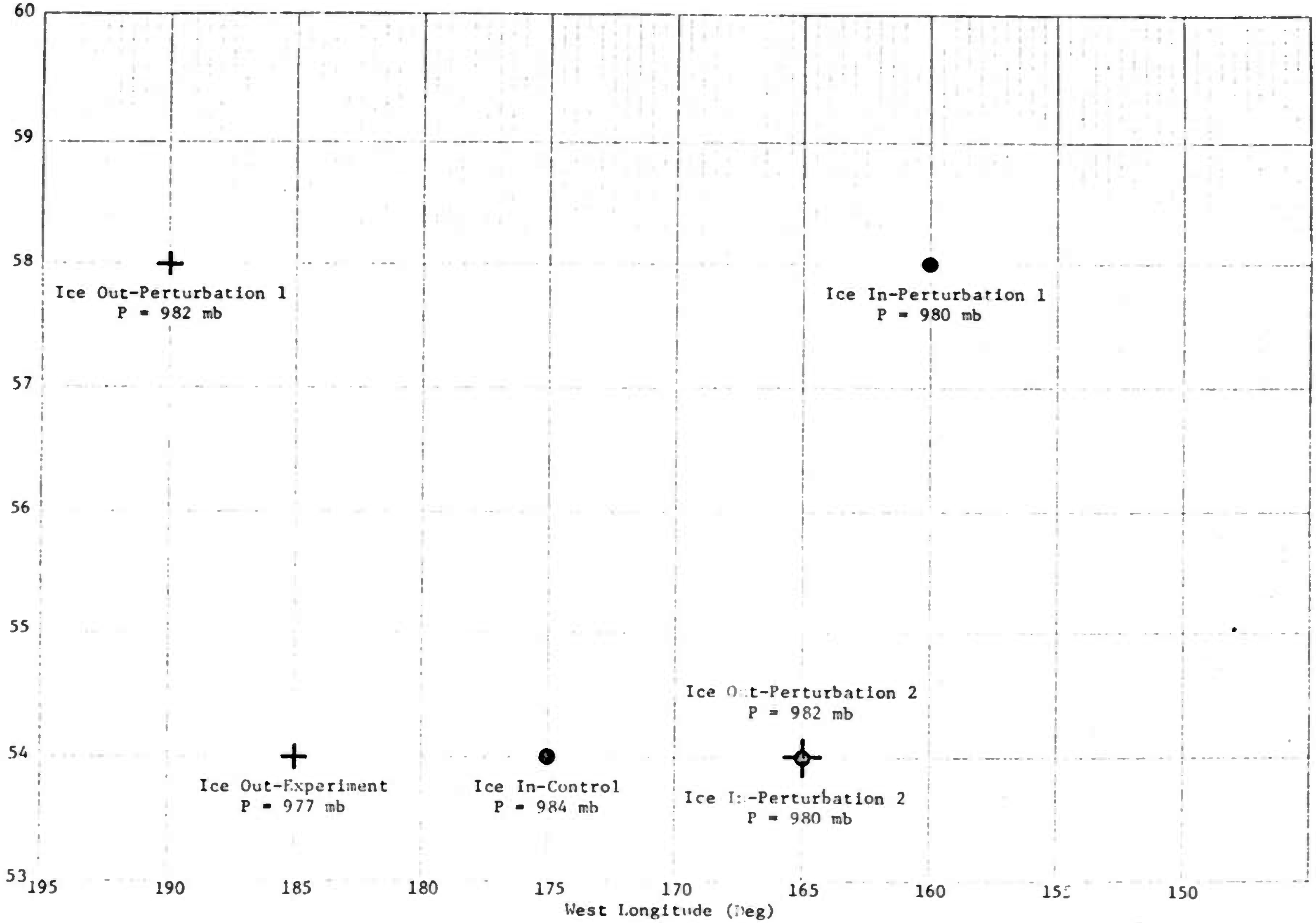

Fig. 1--Location and pressure of the Aleutian low. 30-day mean, Arctic ice-in vs ice-out (3 samples of each).

PRESENT WORK IN HYPOTHESIS TESTING USING TWO SIMULATION RUNS

The central problem in assessing the sensitivity of a climatic simulation is simply how to get enough *independent* samples of both the control and the experiment to permit some reasonably sharp form of hypothesis testing. Instead of making independent runs as described above we are trying to extract equivalent information from the temporal sequence of data generated by a single run.

Remembering that this is not a report on final results, but only an interim statement, we'll outline procedures which are now undergoing tests with both simulated and real data.

Assume the model produces a sequence of data, i.e., a variable of meteorological interest, possibly vector valued, $X_1, X_2, \ldots X_n$. For example, X_i might be the array representing the 800 mb temperatures at every grid point over North America, and the subscript indexes time. For the Mintz-Arakawa model, X_i and X_{i+1} are separated by 6 hours. One procedure is to take all the available data (if the simulation were 60 days long we'd have $X_1, X_2, \ldots, X_{240}$) and find the smallest k such that the subsequence $X_1, X_{1+k}, X_{1+2k}, \ldots, X_{1+nk}$ $(1 + nk \leq 240)$ has the following properties:

- The subsequence has sufficiently small estimated first order correlation coefficient, $\hat{r}_1$. We expect the Durbin-Watson test for absence of first order correlation to provide the appropriate testing procedure. That is, if we accept the null hypothesis from the Durbin-Watson test, we'll claim the subsequence is uncorrelated.
- The subsequence has no significant cyclical component. The turning point test (Kendall, 1963) provides a convenient test of this property. The turning point test counts the number of "peaks" and "troughs" in the time series and tests the hypothesis that this number came from a distribution of independent random variables.

- Last, the sequence should have no significant linear trend. To test for this property, we apply the Kendall rank correlation test (Kendall, 1963). This test takes the sequence X_1, X_2, ... X_n, counts the number of pairs in which $X_j > X_i$, $j > i$. This number, N, is compared with the expected number in a random series and the excess (or deficiency) over the expected number indicates a tendency toward a positive (or negative) trend.

If we fail to reject the null hypothesis of all three tests, we have an adequate demonstration that the random variables in the subsequence X_1, X_{1+k}, ... are uncorrelated and we'll treat them as independent samples. Assuming that there exists a sufficiently small k (or, equivalently, a sufficiently large sample) which still satisfies the above requirements for both the control and the experiment runs, we proceed as follows:

- Using the independent samples from each run, estimate the covariance matrix, G, for the vector variables, X_i. That is, we estimate the covariance between the same meteorological variables at different spatial locations (say, for instance, over the grid points covering North America).
- We ask that the sample size be large enough to permit the inversion of the covariance matrix. Too few samples, while permitting a rough estimation of G, leave G singular and we require G^{-1} to formulate a test statistic.

The investigator, and let's assume he's a meteorologist, is now faced with the first of several interesting questions. They concern the assumptions he's willing to make about the behavior of the real atmosphere in order to achieve a sharper test in discriminating between the control and experimental run. For instance, should he

assume, a priori, that nature would generate identical covariance matrices for identical regions independent of whether the control or experimental condition is being observed? If the answer is yes, we may proceed directly to the final hypothesis testing stage. If the answer is no, the investigator may call for an equality of covariance test (Anderson, 1958) and if the test *rejects* the hypothesis of equality, we have the classical Behrens-Fisher problem (that of testing for the equality of means when the variances are unknown and unequal) for which the hypothesis tests are just not as good as the case of equal covariances. Remember here that a priori knowledge is better than blindly testing for equality, because there is always the non-zero probability of rejecting the hypothesis of equality when, in fact, it is true. Moreover, the equality of covariance test is not very good for large covariance matrices.

In any event, we're finally prepared to use Hotelling's T^2-test (Anderson, 1958) to test the hypothesis that the control and experimental runs have the same mean value, where, again, we mean a vector of mean values of the geographical area of interest.

FILTERING THE DATA AND COVARIANCE MATRIX

In order to get to the T^2-test for equality of two mean vectors, the original data had to meet quite a few criteria. Highly correlated data can only do this at the cost of decreasing sample size. In other words, the more we separate the samples in time trying to achieve independence, the fewer samples we'll have left. There are two procedures which promise to save the really difficult cases:

If we have any reason to believe that the data $\{X_i\}$ are generated by a process which is primarily a Markov process, so that successive values of X are represented by the equation

$$X_i = \rho_1 X_{i-1} + U \qquad (1)$$

where U is a random variable and ρ_1 is the first order correlation coefficient, then the transformation of $\{X_i\}$ to the uncorrelated sequence $\{Y_i\}$ is given by the relationship

$$Y_i = X_i - \hat{\rho}_1 X_{i-1} \tag{2}$$

where $\hat{\rho}_1$ is the first order correlation coefficient *estimated* from the sample $\{X_i\}$.

Now, the closer the sequence $\{X_i\}$ comes to behaving as if it were generated by Eq. (1), the closer $\{Y_i\}$ comes to being an uncorrelated sequence. All good and well except when we examine what effect the transformation has had on the moments. Assume now that we actually have two vector sequence $\{X_i\}$ and $\{Z_i\}$ represented in the sample data from the control and experimental runs. To demonstrate this argument, assume both X_i and Z_j have *identical* mean value, μ, and variance, σ^2. Hence we would expect to *accept* an hypothesis made about the equality of means, since by definition, they're equal. Now, transform $\{X_i\}$ and $\{Z_i\}$ according to

$$Y_i = X_i - \hat{\rho}_x X_{i-1} \tag{3}$$

$$W_i = Z_i - \hat{\rho}_y Z_{i-1} \tag{4}$$

where $\hat{\rho}_x$ and $\hat{\rho}_y$ are the estimated first order correlation coefficients and themselves random variables. The sequences $\{Y_i\}$ and $\{W_i\}$ have the following moments

$$E[Y_i] = (1 - \hat{\rho}_x)\mu \tag{5}$$

$$E[W_i] = (1 - \hat{\rho}_y)\mu \tag{6}$$

$$\mathrm{var}(Y_i) = (1 - \hat{\rho}^2)\sigma^2 \tag{7}$$

$$\mathrm{var}(W_i) = (1 - \hat{\rho}^2)\sigma^2 \tag{8}$$

Notice that the transformation has the favorable property of reducing the variance of each sample, but unfortunately, since $\hat{\rho}_x$ and $\hat{\rho}_y$ are random variables, there is no reason why $E[Y_i]$ and $E[W_i]$ should be equal. In practice, for very large $\hat{\rho}_x$ and $\hat{\rho}_y$, the transformation introduces such a large bias in the means that, even with the variance reducing property, the transformation is not of much value. However, if by testing for the equality of $\hat{\rho}_x$ and $\hat{\rho}_y$ (say, using the Fisher Z-transformation; Hoel, 1956) and finding them equal (or by arguing from the physics that one would *expect* them to be equal -- a risky proposal) we could define a mean correlation coefficient, $\hat{\rho} = \frac{1}{2}(\hat{\rho}_x + \hat{\rho}_y)$, the Markov process transformation could be of great value since it introduces no extra uncertainty into the problem.

Note that even with all these ancillary tests and assumptions, the Markov process transformation doesn't make for sharper discrimination in the final testing of mean values, since the new means are closer together in a way that is not compensated for by the reduction of variance. What we might get, however, is a sufficiently large uncorrelated sample, so at least some hypothesis testing about the equality of means can be done.

Another, even more ad hoc, procedure might be adopted if the sample size is so small that the covariance matrix is singular. If G is an $n \times n$ covariance matrix, then G is non-singular if $T \geq n + 1$ where T is the sample size. In cases where this inequality is not satisfied, we might try formulating the matrix of correlation coefficient, R, associated with G. Then, whenever $r_{ij} < r^*$ set $g_{ij} = 0$, where r^* is some small number. In other words, choose an r^*, say $r^* = 0.1$. Then if the estimated correlation coefficient is less than r^* we arbitrarily

set the corresponding covariance $g_{ij} = 0$. This tends to eliminate small values of g_{ij} which contribute very little to the structure of the problem, but do influence the singularity of G. If this is at all acceptable, the trick will be to find an approximation to the smallest r^* which makes G non-singular. Trial and error might suffice.

MODEL COMPARISONS

There is also the problem of comparing two models or comparing one model to the real atmosphere. Not nearly as much progress has been made here. A trial comparison has been made between the 3-D Mintz-Arakawa model and the zonally averaged model (ZAM) of MacCracken. This was done to test algorithms which produce compatible initial conditions and the actual comparison of predicted results wasn't completed.

The comparison is done by computing the zonal averages of M-A results *after* the run has been made and then comparing these zonal averages with those produced directly by ZAM. Work will accelerate on the comparison task, but we foresee the problem of what constitutes model agreement and the requisite physical explanation should they fail to agree.

The availability of real climatological data should help here. We are preparing a 5-year data base using NCAR raw data. This consists of temperature, geopotential height, wind and relative humidity every 12 hours. We stress the importance of having the sequential data at least daily and of insuring that these data are readily accessible to the investigator and are implementing a data retrieval system which permits easy comparison between the real climate and the M-A results. The same should be true of ZAM data if we make minor variations in our retrieval program, so there's lots that can and will be done in the realm of three-way comparisons.

REFERENCES

Anderson, T. W. (1958): *An Introduction to Multivariate Statistical Analysis*, John Wiley & Sons, New York, 373 pp.

Arakawa, A. (1972): Design of the UCLA General Circulation Model, Technical Report No. 7, *Numerical Simulation of Weather and Climate*, Department of Meteorology, University of California, Los Angeles, 122 pp.

Dyer, R. M. (1971): Method for Filtering Meteorological Data, *Monthly Weather Review*, pp. 435-438.

Fletcher, J. O., et al. (1971): *Numerical Simulation of the Influence of Arctic Sea Ice on Climate*, International Symposium on Physical and Dynamical Climatology, Leningrad, USSR, 16-20 August 1971.

Gates, W. L., et al. (1971): *A Documentation of the Mintz-Arakawa Two-Level Atmospheric General Circulation Model*, R-877-ARPA, The Rand Corporation, Santa Monica, Calif., 408 pp.

Hoel, P. G. (1956): *Introduction to Mathematical Statistics*, John Wiley & Sons, New York, 331 pp.

Kendall, M. G. (1963): *The Advanced Theory of Statistics, Vol. III: Design and Analysis, and Time Series*, Hofner Publishing Company, New York, 552 pp.

Warshaw, M., and R. R. Rapp (1972): *An Experiment on the Sensitivity of a Global Circulation Model: Studies in Climate Dynamics for Environmental Security*, R-908-ARPA, The Rand Corporation, Santa Monica, Calif., 26 pp.